

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

Su-yeon YOON et al.

Application No.: Unassigned

Filing Date: February 24, 2004

Title: CONTINUOUS SPEECH RECOGNITION METHOD AND SYSTEM USING INTER-WORD PHONETIC INFORMATION

Group Art Unit: Unassigned

Examiner: Unassigned

Confirmation No.: Unassigned

SUBMISSION OF CERTIFIED COPY OF PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following priority foreign application(s) in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed.

Country: Korea

Patent Application No(s): 10-2003-0011345

Filed: February 24, 2003

In support of this claim, enclosed is a certified copy(ies) of said foreign application(s). Said prior foreign application(s) is referred to in the oath or declaration. Acknowledgment of receipt of the certified copy(ies) is requested.

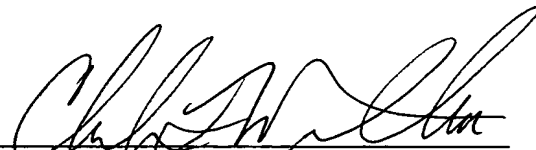
Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620

Date: February 24, 2004

By



Charles F. Wieland III
Registration No. 33,096



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2003-0011345
Application Number

출원 년 월 일 : 2003년 02월 24일
Date of Application FEB 24, 2003

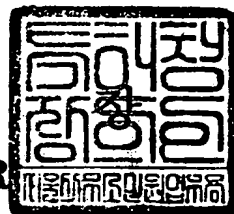
출원인 : 삼성전자주식회사
Applicant(s) SAMSUNG ELECTRONICS CO., LTD.



2003 년 03 월 13 일

특 허 청

COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0007
【제출일자】	2003.02.24
【국제특허분류】	G06F
【발명의 명칭】	음소 결합정보를 이용한 연속 음성인식방법 및 장치
【발명의 영문명칭】	Method and apparatus for speech recognition using phone connection information
【출원인】	
【명칭】	삼성전자 주식회사
【출원인코드】	1-1998-104271-3
【대리인】	
【성명】	이영필
【대리인코드】	9-1998-000334-6
【포괄위임등록번호】	2003-003435-0
【대리인】	
【성명】	이해영
【대리인코드】	9-1999-000227-4
【포괄위임등록번호】	2003-003436-7
【발명자】	
【성명의 국문표기】	윤수연
【성명의 영문표기】	Y00N, Su Yeon
【주민등록번호】	750821-2080316
【우편번호】	135-842
【주소】	서울특별시 강남구 대치3동 대치현대아파트 108동 1101호
【국적】	KR
【발명자】	
【성명의 국문표기】	최인정
【성명의 영문표기】	CHOI, In Jeong
【주민등록번호】	690114-1805019
【우편번호】	441-837

【주소】	경기도 수원시 권선구 권선동 1306 현대아파트 311동 906호
【국적】	KR
【발명자】	
【성명의 국문표기】	김남훈
【성명의 영문표기】	KIM,Nam Hoon
【주민등록번호】	730318-1029714
【우편번호】	442-739
【주소】	경기도 수원시 팔달구 영통동 황골마을주공1단지아파트 101동 1305호
【국적】	KR
【심사청구】	청구
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 이영필 (인) 대리인 이해영 (인)
【수수료】	
【기본출원료】	20 면 29,000 원
【가산출원료】	6 면 6,000 원
【우선권주장료】	0 건 0 원
【심사청구료】	8 항 365,000 원
【합계】	400,000 원
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

음소 결합정보를 이용한 연속 음성인식방법 및 장치가 개시된다. 음소 결합정보를 이용한 연속 음성인식방법은 (a) 각 표제어에 대하여 음운변화가 반영된 발음표기가 수록되며, 표제어간 음운변화에 의하여 상기 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 해당 표제어의 종성 또는 초성에 대한 발음표기에 별도의 식별자를 추가하여 발음사전 데이터베이스를 구축하는 단계; (b) 한국어의 기본음소와 식별자가 추가된 2차적 음소를 이용하여 앞 표제어의 종성에 결합가능한 뒷 표제어의 초성을 대응시켜 가능한 모든 결합으로 이루어진 음소결합정보를 매트릭스화하는 단계; 및 (c) 입력되는 음성으로부터 생성되는 특징벡터에 대하여 상기 발음사전 데이터베이스 및 상기 음소결합정보를 이용하여 음성인식을 수행하는 단계를 포함한다.

【대표도】

도 4

【명세서】**【발명의 명칭】**

음소 결합정보를 이용한 연속 음성인식방법 및 장치 {Method and apparatus for speech recognition using phone connection information}

【도면의 간단한 설명】

도 1은 일반적인 연속 음성인식장치의 구성을 나타낸 블록도,

도 2a 및 도 2b는 도 1에 있어서 발음사전 데이터베이스에 저장된 발음열의 예,

도 3은 도 1에 있어서 탐색부의 동작을 설명하는 도면,

도 4는 본 발명의 일실시예에 따른 연속 음성인식장치의 구성을 나타낸 블록도,

도 5a 및 도 5b는 도 4에 있어서 발음사전 데이터베이스에 저장된 발음열의 예, 및

도 6은 도 4에 있어서 탐색부의 동작을 설명하는 도면이다.

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

<7> 본 발명은 음성인식에 관한 것으로서, 음소 결합정보를 이용한 연속 음성인식방법 및 장치에 관한 것이다.

<8> 일반적으로 연속 음성인식장치는 도 1에 도시된 바와 같은 구성으로 이루어진다. 도 1을 참조하면, 특징추출부(11)에서는 음성인식장치에 입력된 음성을 인식에 유용한 정보만을 추출한 특징벡터로 변환시키고, 탐색부(12)에서는 학습과정에서 미리 구해진 음향모델 데이터베이스(13), 발음사전 데이터베이스(14) 및 언어모델 데이터베이스(15)

를 참조하여 비터비 알고리즘을 이용하여 특징벡터로부터 가장 확률이 높은 단어열을 찾게 된다. 여기서, 대어휘 인식을 위하여 인식대상 어휘들은 트리를 구성하고 있으며, 탐색부(12)에서는 이러한 트리를 탐색한다. 후처리부(16)는 탐색부(12)의 탐색결과로부터 발음기호와 태그 등을 제거하고, 음절 단위로 모아쓰기를 하여 최종 인식결과 즉, 텍스트를 제공한다.

<9> 한편, 발음사전 데이터베이스(14)에 저장된 한국어와 영어에 대한 발음열의 예가 각각 도 2a 및 도 2b에 도시되어 있다. 도 2a를 참조하면, 표제어가 대학(21)인 경우 발생하는 발음열 '대학[dehaKK]'(21a), '대항[deha○○]'(21b), '대하ㄱ[dehag]'(21c) 중 '대항[deha○○]'(21b)과 대항(22)에 대한 발음열 '대항[deha○○]'(22a)이 동일하기 때문에 구분이 불가능하다. 한편, 도 2b를 참조하면, 표제어가 'Seat'(23)인 경우 발생하는 발음열 [sit](23a)과 [sip](23b) 중 [sip](23b)과 'tip'(24)에 대한 발음열 [tip](24a)이 거의 동일하기 때문에 구분이 불가능하다.

<10> 탐색부(12)에서 상기와 같은 발음사전 데이터베이스(14)를 이용하여 탐색과정을 수행하는 예를 도 3을 참조하여 설명하기로 한다. 도 3을 참조하면, '한국대학이'라는 음성을 인식하는 경우, 각각의 어휘 '한국', '대학', '이'에 대하여 초성과 발음이 변하지 않는 몸체 및 종성으로 분리한다. '한국'에 대하여 몸체 부분

에 해당하는 발음열 [aNNgu](32)을 제외한 초성 및 종성으로 결합가능한 발음열(31,33)을 생성하고, '대학'에 대하여 몸체 부분에 해당하는 발음열 [eha](35)을 제외한 초성 및 종성으로 결합가능한 발음열(34,36)을 생성하고, '이'에 대한 발음열(37)을 생성한다. 이와 같이 생성된 발음열에 대하여 Pr(대학 | 한국) 및 Pr(이 | 대학)을 이용하여 탐색 과정을 진행하게 된다. 여기서, '한국'과 '대학' 사이에 결합가능한 경우의 수는 2 가지가 존재하고, '대학'과 '이' 사이에 결합가능한 경우의 수는 3 가지가 존재한다.

<11> 상기한 바와 같이 HMM(Hidden Markov Models) 기반의 연속 음성인식시스템을 구현하기 위해서는 인식 대상 어휘들에 대한 발음열을 정의한 발음사전의 구성이 필수적이다. 통상 음소와 음소 사이 또는 단어와 단어 사이에는 동시조음현상(coarticulation effects)이 빈번하게 발생한다. 이와 같은 동시조음현상이 인식단위의 경계 부분에서 일어나면 각각의 경계를 명확히 구분할 수 없을 뿐 아니라, 동일한 인식단위인 경우에도 앞뒤 인식단위 문맥에 따라 다른 음향학적 특성을 갖기도 하므로 인식단위 모델링에 있어서는 이러한 현상들이 잘 반영되어야 한다. 특히, 한국어 음성언어의 경우 다양한 음소문맥에 따라 많은 음운변화 현상이 나타나기 때문에, 각 발음열은 이러한 음운변화 현상을 잘 반영하고 있어야 한다. 그러나, 표제어내의 발음열들은 비교적 고정된 음소문맥을 가지고 있어서 트라이폰과 같은 인식단위 학습을 통한 음운변화 현상 모델링이 용이한 반면, 표제어와 표제어 사이에서는 인접한 표제어에 따라 다양한 음소문맥이 가능하기 때문에 이러한 음운변화 현상을 반영하려면 보다 정교한 모델링 방법이 제시되어야 한다.

<12> 표제어간의 음운변화 현상을 반영하기 위하여 첫째, 표제어 경계에서 나타날 수 있는 모든 가능한 음소문맥 또는 주로 나타나는 음소문맥을 포함하여 사전에 다중 발음열

형태로 수록하는 방법을 사용하거나 둘째, 보다 많은 수의 혼합 가우시안 함수를 이용하여 좀 더 정교한 HMM의 상태 출력값을 통해 변화 현상을 모델링하는 방법을 사용할 수 있다. 첫째 방법은, 사전 및 네트워크 사이즈가 너무 방대해 지고 두번째 방법은 계산량이 많아짐으로써, 인식에 소요되는 시간이 길어지게 되어 인식속도가 늦어지는 단점이 있다. 셋째, 표제어간에 주로 나타나는 음운변화 현상을 선정한 후, 언어모델이 고려된 변형된 음소문맥을 단어간 트라이폰(crossword triphone)을 이용하여 인식 네트워크에 직접 적용하는 방법을 사용할 수 있는데, 이 경우 단어 앞 부분에 각각 다른 음소문맥을 반영하였기 때문에 동일한 단어라도 다른 시작노드에서 출발하게 되고, 그 결과 트리구조를 기반으로 하는 인식 네트워크에서는 공유효율이 떨어지고 네트워크의 크기가 방대해 지는 단점이 있다. 넷째, 트리 구조의 인식 네트워크에서 단어 앞 부분에 대한 음소문맥을 미리 인식 전에 적용시키지 않고 인식중에 적용시키는 방법을 사용할 경우, 같은 음운환경에서 2가지로 변화하는 경우 이를 제한할 수 없을 뿐 아니라 인식과정에서 매 프레임에 대해 발음규칙을 적용해야 하므로 부담이 증가하고, 인식과정 중 인식 네트워크를 계속 업데이트시켜야 하는 부담이 큰 단점이 있다.

【발명이 이루고자 하는 기술적 과제】

<13> 따라서 본 발명이 이루고자 하는 기술적 과제는 표제어의 철자와 발음이 동일하지 않은 경우 식별자를 이용한 음소결합정보를 이용하여, 연속어 음성인식과정에서 발생하는 표제어간 음운규칙의 부적합한 발음열들의 결합을 제거하는 연속 음성인식방법을 제공하는데 있다.

<14> 본 발명이 이루고자 하는 다른 기술적 과제는 상기 음소 결합정보를 이용한 연속 음성인식방법을 실현하는데 가장 적합한 장치를 제공하는데 있다.

<15> 상기 기술적 과제를 달성하기 위하여 본 발명에 따른 음소 결합정보를 이용한 연속 음성인식방법은 (a) 각 표제어에 대하여 음운변화가 반영된 발음표기가 수록되며, 표제어간 음운변화에 의하여 상기 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 해당 표제어의 종성 또는 초성에 대한 발음표기에 별도의 식별자를 추가하여 발음사전 데이터베이스를 구축하는 단계; (b) 한 국어의 기본음소와 식별자가 추가된 2차적 음소를 이용하여 앞 표제어의 종성에 결합가능한 뒷 표제어의 초성을 대응시켜 가능한 모든 결합으로 이루어진 음소결합정보를 매트릭스화하는 단계; 및 (c) 입력되는 음성으로부터 생성되는 특징벡터에 대하여 상기 발음사전 데이터베이스 및 상기 음소결합정보를 이용하여 음성인식을 수행하는 단계를 포함한다.

<16> 상기 (c) 단계는 바람직하게로는 상기 음소 결합정보를 이용하여 앞 표제어에서 마지막 음절의 종성에 대한 발음열과 뒷 표제어에서 첫 음절의 초성에 대한 발음열 간의 연결을 제한하는 단계를 포함한다.

<17> 상기 다른 기술적 과제를 달성하기 위하여 본 발명에 따른 음소 결합정보를 이용한 연속 음성인식장치는 학습에 의해 구축된 음향모델 데이터베이스 및 언어모델 데이터베이스를 포함하는 연속 음성인식장치에 있어서, 표제어간 음운변화로 인하여 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 초성 또는 종성에 식별자를 추가하고, 기본 음소와 상기 식별자가 추가된 2차 음소를 이용하여 상기 앞 표제어에서 마지막 음절의 종성에 결합가능한 뒷 표제어에서 첫 음절의 초성을 대응시킨 가능한 모든 결합으로 이루어진 음소결합정보를 데이터베이스화하여 저장하는 음소 결합정보 저장부; 각 표제어에 대하여 음운변화가 반영된 발음

표기가 수록되며, 표제어간 음운변화에 의하여 상기 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 해당 표제어의 종성 또는 초성에 대한 발음표기에 별도의 식별자를 추가하여 구축된 발음사전 데이터베이스; 입력되는 음성으로부터 인식에 유용한 정보들을 추출하여 특징벡터로 변환하는 특징추출부; 및 상기 음소 결합정보를 이용하고 상기 음향모델 데이터베이스, 상기 발음사전 데이터베이스 및 상기 언어모델 데이터베이스를 참조하여, 상기 특징추출부에서 변환된 특징벡터와 유사할 확률이 가장 높은 단어열을 탐색하고, 탐색된 단어열을 인식된 텍스트로서 출력하는 탐색부를 포함한다.

<18> 한편, 상기 장치는 바람직하게로는 상기 탐색부에서 탐색된 단어열에 대하여 상기 음소 결합정보를 이용하고, 상기 음향모델 데이터베이스에서 표제어간 트라이폰 모델을 참조하여 리스코어링을 수행하는 후처리부를 더 포함한다.

【발명의 구성 및 작용】

<19> 이하, 본 발명의 실시예에 대하여 첨부된 도면들을 참조하여 상세하게 설명하기로 한다.

<20> 도 4는 본 발명의 일실시예에 따른 연속 음성인식장치의 구성을 나타낸 블럭도로서, 특징추출부(41), 탐색부(42), 음향모델 데이터베이스(43), 발음사전 데이터베이스(44), 음소 결합정보 저장부(45), 언어모델 데이터베이스(46) 및 후처리부(47)로 이루어진다.

<21> 도 4를 참조하면, 특징추출부(41)에서는 환경적응, 끝점검출, 반향제거 또는 잡음제거 등과 같은 전처리과정을 거친 다음 디지털처리된 음성신호를 효과적으로 표현해 주

는 특징벡터를 추출한다. 특징벡터 추출에는 일반적으로 특징벡터의 저차항에는 발화 당시의 성도(vocal tract) 특성을 반영하고, 고차항에는 발화를 이끌게 한 기저(excitation) 신호의 특성이 반영되도록 한 캡스트럼(cepstrum) 추출방식이 사용되고, 최근에는 인간의 청각인지과정을 반영한 캡스트럼 추출방식인 MFCC(Mel Frequency Cepstrum Coefficient)가 사용되기도 한다.

<22> 탐색부(42)에서는 단어열의 탐색을 위하여 후술할 음소 결합정보 저장부(45)에 저장된 음소 결합정보를 이용하고, 학습과정에서 미리 구해진 음향모델 데이터베이스(43), 발음사전 데이터베이스(44) 및 언어모델 데이터베이스(46)를 참조하여 비터비 알고리즘을 이용하여 특징벡터로부터 가장 확률이 높은 단어열을 찾게 된다. 여기서, 대어휘 인식을 위하여 인식대상 어휘들은 트리를 구성하고 있으며, 탐색부(42)에서는 음향모델 데이터베이스(43)의 음향학적인 대비결과에 대응하는 언어모델 데이터베이스(46)의 발생빈도 데이터 및 발생확률 데이터를 참조하여 발음사전 데이터베이스(44)에 수록된 단어들을 조합하여 트리를 탐색함으로써 N개의 베스트 후보들을 구성한다.

<23> 음향모델 데이터베이스(43)는 학습용 발성 데이터베이스의 음성데이터로부터 추출된 사용자가 갖는 특징벡터들을 이용하여 은닉 마르코브 모델(Hidden Markov Models) 등과 같은 음향모델들을 형성한다. 이들 음향모델들은 음성인식과정에서 기준 모델로서 이용된다.

<24> 발음사전 데이터베이스(44)는 음운변화가 적용된 인식단위를 처리해야 하므로 각 표제어에 대하여 음운변화 현상이 반영된 발음표기를 모두 수록하여 데이터베이스화한 것이다. 이때, 표제어간 음운변화에 의하여 앞 표제어에서 마지막 음절의 종성이나 뒷

표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 해당 표제어의 종성 또는 초성에 대한 발음표기에 별도의 식별자를 추가하여 수록하게 된다.

<25> 음소결합정보 저장부(45)는 표제어간 결합에서 결합할 수 있는 모든 음소간의 연쇄가 매트릭스화하여 저장되며, 기본 음소와 식별자가 추가된 2차 음소를 이용하여 앞 표제어에서 마지막 음절의 종성에 결합가능한 뒷 표제어에서 첫 음절의 초성을 대응시킨 가능한 모든 결합으로 이루어진 음소결합정보를 저장한다. 표제어간의 음운변화 현상은 앞 표제어에서 마지막 음절의 종성과 뒷 표제어에서 첫 음절의 초성의 쌍으로 나타낼 수 있는데, 모든 쌍에서 음운변화가 일어나는 것이 아니라 일정한 규칙에 따라 특정한 쌍에서 일어나게 된다. 즉, 복합어 경계 및 어절간 음소경계 변환규칙들을 이용하여 변환이 일어나는 쌍에서 각 음소의 변화를 알아낸다면 상당히 정확하게 단어간 음운변화 현상을 모델링할 수 있다. 음소 결합정보를 발생시키는 표제어 간에 발생하는 음운변화 현상의 유형은 대치, 삽입, 삭제로 분류할 수 있다. 대치 현상은 장애음/유음의 비음화 현상이나, 변자음화 현상, 경음화와 같은 규칙들에 의해, 인접한 두 자음 모두 또는 한 쪽이 다른 자음으로 변형된다. 삽입현상은 주로 모음으로 시작하며, 연음법칙이 적용되는 경우에 뒷단어에서 일어나거나 'ㄴ' 첨가현상에 의해 일어난다. 삭제 현상은 대체적으로 삽입현상과 함께 나타나며, 삽입 현상이 뒷 단어의 초성에 일어나는 반면, 삭제 현상은 앞 단어의 종성에서 일어난다. 상기의 음운변화현상을 거쳐 변형된 음소는 식별자를 추가하여 변형되지 않은 음소와 구별하며, 앞 표제어에서 마지막 음절의 종성에 결합가능한 뒷 표제어의 첫 음절의 초성을 대응시킨 모든 음소결합정보가 데이터베이스화된다. 음소 결합정보 저장부(45)에 저장된 음소 결합정보는 탐색부(42)에서의 탐색과정

및 후처리부(47)에서의 리스코어링 과정에 적용함으로써 탐색과정 및 리스코어링 과정에서 탐색후보를 줄일 수 있기 때문에 탐색과정을 보다 정밀화할 수 있다.

<26> 언어모델 데이터베이스(46)는 학습용 텍스트 데이터베이스에 구축된 학습용 문장들에서의 단어열에 대한 발생빈도 데이터를 이용하여 연산되어진 단어열들의 바이그램 또는 트라이그램 발생확률 데이터를 구비한다. 여기서, 학습용 텍스트 데이터베이스는 음성인식에 사용될 가능성이 있는 문장들로 이루어지며, 이는 통상 제작자에 의해 사용될 가능성이 있다고 판단된 것들로서 그 양이 한정된다.

<27> 후처리부(47)는 탐색부(42)에서 탐색된 단어열에 대하여 음소 결합정보 저장부(45)에 저장된 음소 결합정보와, 음향모델 데이터베이스(43), 발음사전 데이터베이스(445) 및 언어모델 데이터베이스(46)를 참조하여 리스코어링(rescoring)을 수행하고, 최종 리스코어링된 단어열로부터 발음기호와 태그 등을 제거하고, 음절 단위로 모아쓰기를 하여 최종 인식결과 즉, 텍스트를 제공한다. 즉, 리스코어링과정에서는 후방향 탐색과정에서 얻어진 최적의 발음열에 대하여, 표제어간 경계에서 표제어내의 바이폰 모델을 표제어간 트라이폰 모델로 확장, 변형시킨 후, 각 후보별 문장점수를 재계산하고, 최종적인 인식결과를 선정한다. 이와 같이, 가장 확률이 높은 몇개의 후보문장에 대해서만 표제어간 경계에서는 바이폰 모델을 사용하던 것을 트라이폰 모델로 변형시킨 후 음향모델 확률값을 재계산하면, 인식효율을 유지시키면 인식성능을 향상시킬 수 있다.

<28> 여기서, 특징추출부(41), 탐색부(42) 및 후처리부(47)는 연산기능을 가진 프로세서, 예를 들면 CPU(Central Processing Unit)와 그에 의해 구동되는 음성인식 프로그램에 의해 구현될 수 있으며, 음향모델 데이터베이스(43), 음소 결합정보 저장부(44), 발음사

전 데이터베이스(45) 및 양방향 언어모델 데이터베이스(46)는 메모리에 의해 구현될 수 있다.

<29> 한편, 음소 결합정보 저장부(44)에 저장되는 표제어간 음소결합정보의 예를 나열하면 다음 표 1에 도시된 바와 같다.

<30> 【표 1】

앞 표제어 마지막음절 종성(C)	뒷 표제어 첫음절 초성(D)	경계 정보
00(○)	n(ㄴ)	0
00(○)	n(ㄴ)	1
00(○)	n(ㄴ)	2
00(○)	m(ㅁ)	0
00(○)	m(ㅁ)	1
00(○)	m(ㅁ)	2
00(○)	b(ㅂ)	0
00(○)	b(ㅂ)	1
00(○)	b(ㅂ)	2
00(○)	s(ㅅ)	0
00(○)	s(ㅅ)	1
00(○)	s(ㅅ)	2
KK(ㄱ)	b1(1)(ㅂ→ㅅ)	0
KK(ㄱ)	b1(1)(ㅂ→ㅅ)	1
KK(ㄱ)	b(ㅂ)	1
KK(ㄱ)	b(ㅂ)	2
KK(ㄱ)	s1(1)(ㅅ→ㅅ)	0
KK(ㄱ)	s1(1)(ㅅ→ㅅ)	1
KK(ㄱ)	s(ㅅ)	1
KK(ㄱ)	s(ㅅ)	2
00(1)(ㄱ→○)	n(ㄴ)	0
00(1)(ㄱ→○)	m(ㅁ)	0

<31> 상기 표 1에 도시된 바와 같이, 어휘간의 음운변화 현상에 따르면 앞 표제어에서 마지막음절의 종성 또는 뒷 표제어에서 첫음절의 초성의 철자와 이에 대한 발음표기가 달라지는 경우가 한정되어 있기 때문에 앞 표제어에서 마지막음절의 종성 또는 뒷 표제

어에서 첫음절의 초성의 철자와 이에 대한 발음표기가 달라지도록 하는 뒷 표제어에서 첫음절의 초성 또는 앞 표제어에서 마지막음절의 종성을 대응시켜 음소결합정보로서 음소결합정보 저장부(44)에 저장한다. 예를 들면, 앞 표제어에서 마지막 음절의 종성이 'ㄱ'인 경우 뒷 표제어에서 첫음절의 초성이 'ㄴ' 또는 'ㄹ'이 되는 경우 앞 표제어의 철자와 발음표기가 달라지므로, 앞 표제어의 종성 발음열 [00(1)]는 뒷 표제어의 초성 발음열 [n] 또는 [m]로만 결합되도록 한다. 즉, 음소결합정보로는 "ㄱㄴ→[00(1)n]"이 아니라 [00(1)n]만 저장되며, 발음사전 데이터베이스(44)에도 식별자가 이미 추가되어 있으므로 [00(1)n] 정보만으로 결합불가능한 경우를 제약할 수 있는 것이다.

<32> 한편, 표 1에 있어서 경계정보는 문장발성시 쉼(pause)의 유무를 표시하는 것이다. 사람은 문장을 발성하는 도중에 숨을 쉬게 되며, 이 경우 앞 표제어와 뒷 표제어간에 음운변환이 발생하지 않는다. 경계정보는 문장발성시의 쉼의 강도에 따라서 0(쉬지 않는 경우), 1(약하게 쉬는 경우), 2(강하게 쉬는 경우)로 구분한 것으로, 경계정보에 따라서 앞 표제어 마지막 음절의 종성에 결합할 수 있는 뒷 표제어 첫음절 초성이 달라지게 된다. 경계정보에 따라 구분하여 음소결합정보 저장부(45)에 저장한다.

<33> 다음, 발음사전 데이터베이스(44)에 저장된 한국어와 영어에 대한 발음열의 예를 각각 도 5a 및 도 5b를 참조하여 설명하기로 한다. 도 5a를 참조하면, 표제어가 대학(51)인 경우 '대학[dehaKK]'(51a), '대항[deha○○(1)]'(51b), '대하ㄱ[dehag]'(51c)이라는 3가지 발음열이 존재하고, 표제어가 대항(22)인 경우 '대항[deha○○]'(52a)이라는 발음열이 존재한다. 이때, 표제어인 대학(51)과 그 발음표기가 동일하지 않은 발음열 '대항[deha○○(1)]'(51b)에 대해서는 종성 발음열의 후단에 별도의 식별자 예를 들면 '(1)'이 부가된다. 따라서, 대학(51)에 대하여 음운변화로 발생하는 발음열 '대항[deha

○○(1)']'(51b)과 대항(52)에 대한 발음열 '대항[deha○○]'(52a)에 대하여 구분이 가능해진다. 이와 같이 식별자가 표기된 앞 표제어에서 마지막 음절의 종성에 대한 발음열은 상기 표 1에서와 같이 뒷표제어의 첫음절 초성에 특정한 발음열이 따라 오는 경우에만 앞 표제어와 뒷 표제어간의 결합이 가능해진다.

<34> 도 5b를 참조하면, 표제어가 'Seat'(53)인 경우 [sit](53a)과 [sip](53b)이라는 2가지 발음열이 존재하고, 표제어가 'tip'(54)인 경우 [tip](54a)이라는 발음열이 존재한다. 이때, 표제어인 'Seat'(53)와 그 발음표기가 동일하지 않은 발음열 [sip(1)](53b)에 대해서는 도 5a에서와 마찬가지로 발음열 후단에 별도의 식별자 예를 들면 '(1)'이 부가된다. 따라서, 'Seat'(53)에 대하여 음운변화로 발생하는 발음열 [sip(1)](53b)과 'tip'(54)에 대한 발음열 [tip](54a)에 대하여 구분이 가능해진다. 즉, 발음열 [sip(1)](53b)는 뒷 표제어의 음소가 'b', 'p', 'm'인 경우에만 결합가능하기 때문에 예를 들면 'down'과의 결합 자체가 미리 차단될 수 있다. 한편, 발음열 [tip](54a)은 뒷 표제어의 음소에 상관없이 결합가능하다.

<35> 도 6은 도 4에 도시된 탐색부(42)에서 발음사전 데이터베이스(44)와 음소 결합정보 저장부(45)를 이용하여 탐색과정을 수행하는 예를 설명하는 것이다. 도 6을 참조하면, '한국대학이'라는 음성을 인식하는 경우, 먼저 소정의 인식단위, 여기서는 표제어 즉 '한국', '대학', '이'로 분리하여 각 표제어에 대한 발음열 네트워크를 진행하여 최적의 발음열이 선택된다. 발음열 네트워크에는 음운변화 규칙이 적용되어 발생가능한 모든 발음열들로 트리가 구성되어 있다. 이와 같이 선택된 발음열에 대하여 Pr(대학 | 한국) 및 Pr(이 | 대학) 및 음소결합정보 저장부(45)로부터 제공되는 음소결합정보를 이용하여 탐색과정을 진행하게 된다. 음소 결합정보를 이용함으로써 '한국'의 종성에 대한 발음

열 [00(1)]과 '대학'의 초성에 대한 발음열 [d]의 결합 자체가 차단되고, '대학'의 종성에 대한 발음열 [00(1)]과 '이'의 초성에 대한 발음열[i]의 결합과 '대학'의 종성에 대한 발음열 [k(1)]과 '이'의 초성에 대한 발음열[i]의 결합 자체가 차단되어, 전체적인 탐색 네트워크가 단순해질 수 있다. 즉, 탐색부(42)에서 전방향 탐색 및 후방향 탐색과정을 수행함에 있어서, 앞 표제어에서 뒤 표제어로 전이될 때마다 후속하는 표제어의 다수의 발음열 중, 앞 표제어와 결합가능한지 여부를 음소결합정보를 이용하여 확인한 후, 부적절한 결합을 미리 제거한다. 특히, 후방향 탐색과정은 전방향 탐색과정에서 선정된 각 프레임에 대한 인식후보단어를 이용하여 문장을 구성하는 과정으로서, 이때 언어모델 확률값을 이용하여 표제어간 가장 적합하게 연결된 소정 갯수, 예를 들면 10개의 후보문장을 구성한다. 리스코어링 단계에서는 후방향 탐색과정에서 선정된 10개의 후보문장에 대해 표제어 경계에서 사용한 표제어 내부의 바이폰 모델을 표제어간 트라이폰 모델로 변환하고, 음소결합정보와 음향모델을 이용하여 확률값을 재계산하여 최적의 인식 문장을 찾아낸다.

<36> <실험예>

<37> 다음, 본 발명에 따른 음성인식방법과 종래의 음성인식방법의 성능을 비교하기 위한 실험을 하였다. 주변 잡음이 없는 조용한 사무실 환경에서 화자 600명이 발화한 한국어 낭독체 45000 문장을 학습데이터로, 조용한 사무실 환경에서 화자 80명이 발화한 한국어 낭독체 480 문장을 평가데이터로 사용하였다. 실험에 사용된 특징벡터는 12차의 MFCC(Mel Frequency Cepstral Coefficients) 계수와 에너지, 그리고 이들의 차분계수로서, 한 프레임이 26차의 벡터로 표현된다. 실험을 위하

여 44 베이스폰의 4000 서브워드 모델로 이루어지며, PTM(Phonetically-tied Mixture) 모델 구조를 가지는 음향모델이 사용되었으며, 트라이그램 언어모델이 사용되었다. 사용된 어휘는 정치면 기사와 관련된 11000 단어로서, 의사형태소 단위로 인식된다. 상기와 같은 조건에 대한 실험결과는 다음 표 2에 도시된 바와 같다.

<38> 【표 2】

인식방법			단어인식율(%)	문장인식율(%)
발음사전	리스크어링	경계정보	에러율(%)	에러율(%)
기존 발음사전	X	X	10.74	53.96
기존 발음사전	0	X	9.03	48.33
본 발명의 발음사전	X	X	8.87	47.50
본 발명의 발음사전	0	X	7.79	42.92
본 발명의 발음사전	X	0	7.58	41.27
본 발명의 발음사전	0	0	6.53	40.00

<39> 상기 표 2를 살펴보면, 단어인식률 및 문장인식율에 있어서 본 발명에 의한 발음사전을 이용한 경우 기존 발음사전을 이용했을 경우에 비하여 성능향상을 보이며, 그 중에서도 리스크어링 단계와 경계정보를 이용한 경우 가장 우수한 성능을 가지는 것이 입증되었다.

<40> 상기한 본 발명의 실시예는 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 한편, 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식

으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고 본 발명을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있다.

【발명의 효과】

- <41> 상술한 바와 같이 본 발명에 따르면, 서브워드의 증가나 음향모델의 변화없이 적용 가능하며, 음소 결합정보를 이용하여 부적합한 표제어간 결합을 제거할 수 있으므로 인식율을 향상시킬 수 있고, 인식 도중에 음운규칙을 사용하지 않기 때문에 디코딩 과정에 부담을 주지 않는 이점이 있다.
- <42> 또한, 어휘별로 문맥 정보를 기입하지 않으므로 발음사전 데이터베이스의 크기 및 탐색 네트워크가 증가하지 않을 뿐 아니라, 언어에 종속적인 음운규칙을 구성하는 방법과는 달리 발음을 표제어의 철자와의 동일여부에 따라 세분화하고, 이의 결합관계를 자동으로 추출함으로써 한국어 뿐만 아니라 외국어에도 적용할 수 있는 이점이 있다.
- <43> 이상 도면과 명세서에서 최적 실시예들이 개시되었다. 여기서 특정한 용어들이 사용되었으나, 이는 단지 본 발명을 설명하기 위한 목적에서 사용된 것이지 의미 한정이나 특허청구범위에 기재된 본 발명의 범위를 제한하기 위하여 사용된 것은 아니다. 그러므로 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의해 정해져야 할 것이다.

【특허청구범위】**【청구항 1】**

(a) 각 표제어에 대하여 음운변화가 반영된 발음표기가 수록되며, 표제어간 음운변화에 의하여 상기 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 해당 표제어의 종성 또는 초성에 대한 발음표기에 별도의 식별자를 추가하여 발음사전 데이터베이스를 구축하는 단계;

(b) 한국어의 기본음소와 식별자가 추가된 2차적 음소를 이용하여 앞 표제어의 종성에 결합가능한 뒷 표제어의 초성을 대응시켜 가능한 모든 결합으로 이루어진 음소결합정보를 매트릭스화하는 단계; 및

(c) 입력되는 음성으로부터 생성되는 특징벡터에 대하여 상기 발음사전 데이터베이스 및 상기 음소결합정보를 이용하여 음성인식을 수행하는 단계를 포함하는 연속 음성인식방법.

【청구항 2】

제1 항에 있어서, 상기 (c) 단계는 상기 음소결합정보를 이용하여 앞 표제어에서 마지막 음절의 종성에 대한 발음열과 뒷 표제어에서 첫 음절의 초성에 대한 발음열 간의 연결을 제한하는 단계를 포함하는 연속 음성인식방법.

【청구항 3】

제1 항에 있어서, 상기 음소결합정보는 경계정보에 따라 구분하여 매트릭스화되는 것을 특징으로 하는 연속 음성인식방법.

【청구항 4】

제1 항 내지 제3 항 중 어느 한 항에 기재된 방법을 실행시킬 수 있는 프로그램을 수록한 컴퓨터로 판독가능한 기록매체.

【청구항 5】

학습에 의해 구축된 음향모델 데이터베이스 및 언어모델 데이터베이스를 포함하는 연속 음성인식장치에 있어서,

표제어간 음운변화로 인하여 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우, 초성 또는 종성에 식별자를 추가하고, 기본 음소와 상기 식별자가 추가된 2차 음소를 이용하여 상기 앞 표제어에서 마지막 음절의 종성에 결합가능한 뒷 표제어에서 첫 음절의 초성을 대응시킨 가능한 모든 결합으로 이루어진 음소결합정보를 데이터베이스화하여 저장하는 음소 결합정보 저장부;

각 표제어에 대하여 음운변화가 반영된 발음표기가 수록되며, 표제어간 음운변화에 의하여 상기 앞 표제어에서 마지막 음절의 종성이나 뒷 표제어에서 첫 음절의 초성의 철자와 발음이 동일하지 않은 경우 해당 표제어의 종성 또는 초성에 대한 발음표기에 별도의 식별자를 추가하여 구축된 발음사전 데이터베이스;

입력되는 음성으로부터 인식에 유용한 정보들을 추출하여 특징벡터로 변환하는 특징추출부; 및

상기 음소 결합정보를 이용하고 상기 음향모델 데이터베이스, 상기 발음사전 데이터베이스 및 상기 언어모델 데이터베이스를 참조하여, 상기 특징추출부에서 변환된 특징

벡터와 유사할 확률이 가장 높은 단어열을 탐색하고, 탐색된 단어열을 인식된 텍스트로서 출력하는 탐색부를 포함하는 연속 음성인식장치.

【청구항 6】

제5 항에 있어서, 제1 항에 있어서, 상기 음소결합정보는 경계정보에 따라 구분하여 매트릭스화되는 것을 특징으로 하는 연속 음성인식장치.

【청구항 7】

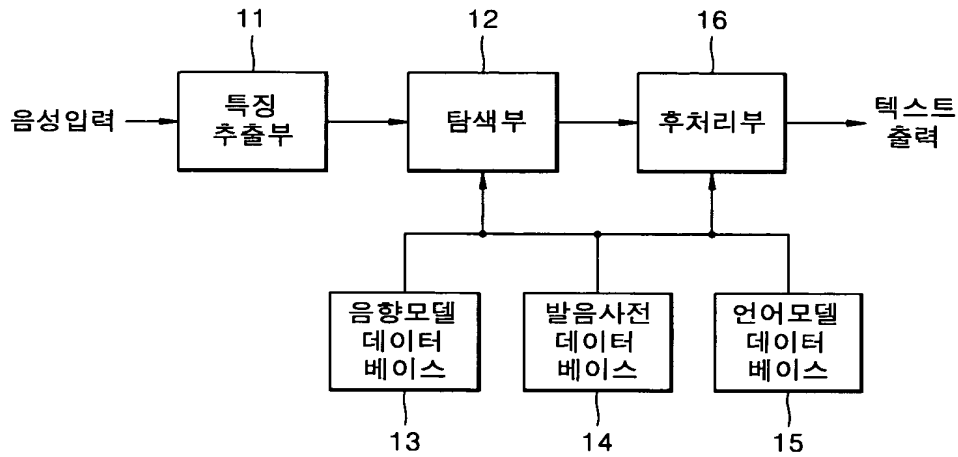
제5 항에 있어서, 상기 탐색부는 상기 음소 결합정보를 이용하여 앞 표제어의 종성에 대한 발음열과 뒷 표제어의 초성에 대한 발음열 간의 연결을 제한하는 것을 특징으로 하는 연속 음성인식장치.

【청구항 8】

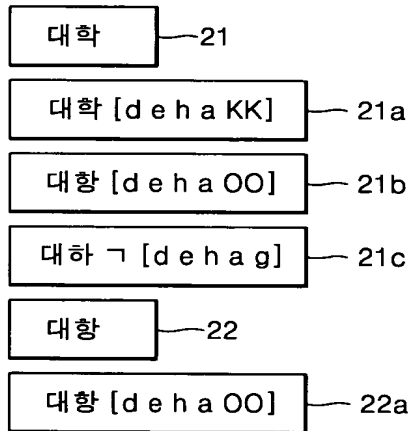
제5 항에 있어서, 상기 장치는 상기 탐색부에서 얻어진 최적 발음열에 대하여, 표제어간 바이폰 모델을 트라이폰 모델로 변환한 후, 트라이폰 모델을 참조하여 음향모델 스코어를 리스코어링하여 각 후보별 문장점수를 재계산하고 최종적인 인식결과를 선정하는 후처리부를 더 포함하는 연속 음성인식장치.

【도면】

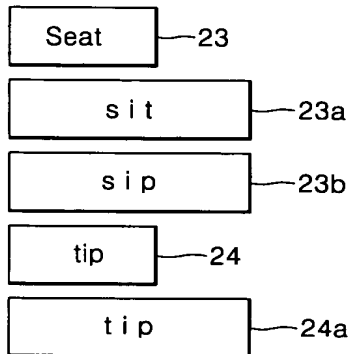
【도 1】



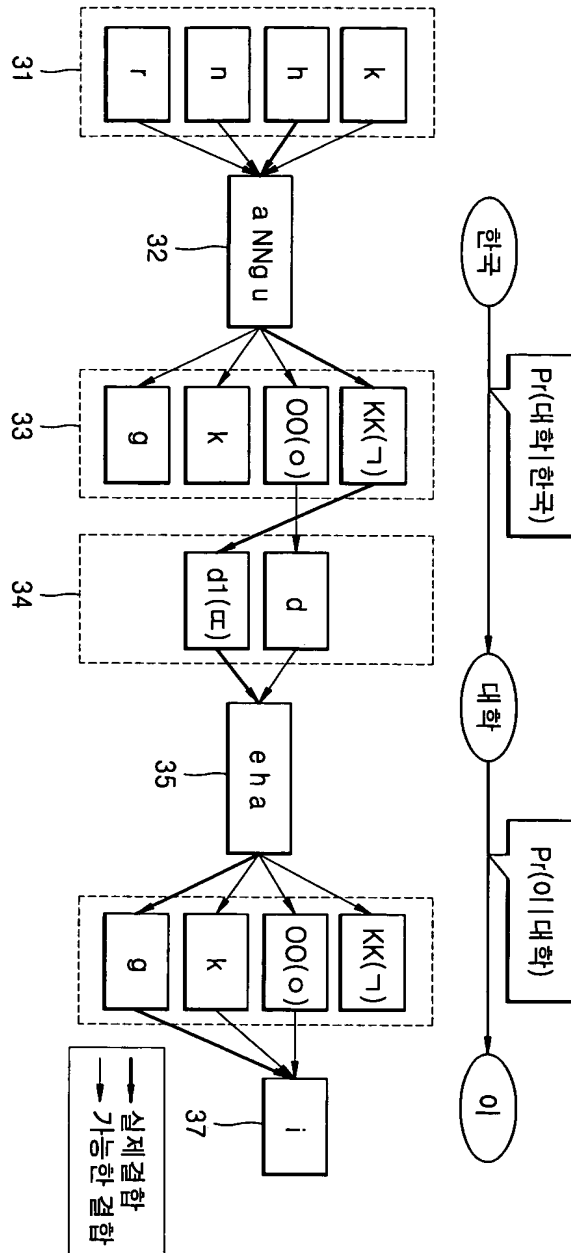
【도 2a】



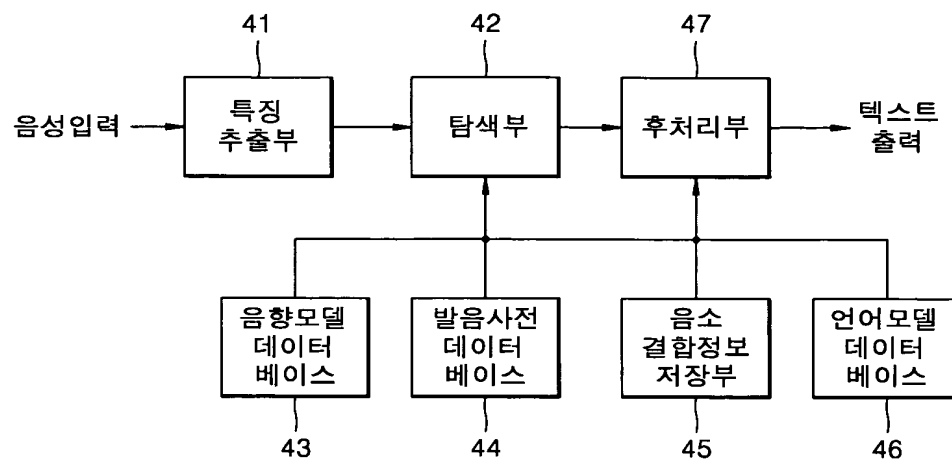
【도 2b】



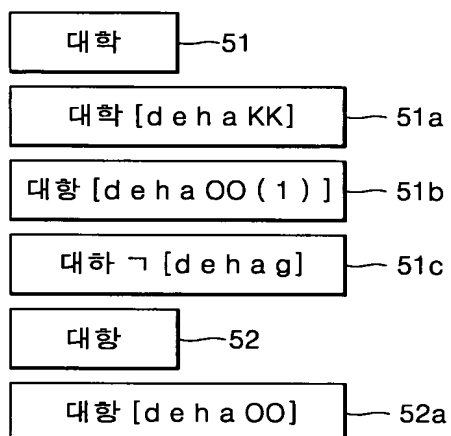
【도 3】



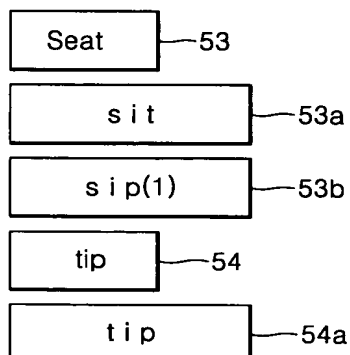
【도 4】



【도 5a】



【도 5b】



【도 6】

